DTIC FILE COPY

AD-A210 273

[redacted]

PURDUE UNIVERSITY

DTIC
ELECTE
JUL 1 4 1989
S D
D

CENTER FOR STATISTICAL DECISION SCIENCES AND DEPARTMENT OF STATISTICS

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

89 7 13 05

(1)

EMPIRICAL BAYES SELECTION FOR THE HIGHEST PROBABILITY OF SUCCESS IN NEGATIVE BINOMIAL DISTRIBUTIONS *

by

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, MI 48202

Technical Report #89-18C

Department of Statistics
Purdue University

June 1989

* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620 at Purdue University.

★

EMPIRICAL BAYES SELECTION FOR THE HIGHEST PROBABILITY OF SUCCESS IN NEGATIVE BINOMIAL DISTRIBUTIONS

by

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, MI 48202

ABSTRACT

We study the problem of selecting the highest probability of success from among several negative binomial distributions via the nonparametric empirical Bayes approach. A monotone selection rule is proposed on basis of monotone empirical Bayes estimators of the negative binomial success probabilities which are obtained by using the antitonic and isotonic regression methods. The asymptotic optimality property of the proposed empirical Bayes selection rule is also established.

Key Words and Phrases: Asymptotically optimal; empirical Bayes; monotone estimator; monotone selection rule; isotonic and antitonic regression.

1. INTRODUCTION

In many situations, an experimenter is often confronted with choosing a model which is the best in some sense among those under study. For example, consider k different competing drugs for a certain ailment. One would like to select the best among them in the sense that it has the highest probability of success (cure of the ailment). This kind of selection problem occurs in many fields, such as medicine, engineering, and sociology. The reader is referred to Gupta and Panchapakesan (1979) for further discussions on goals and procedures for this selection problem.

Now, consider a situation in which one will be repeatedly dealing with the same selection problem independently. In such instances, it is reasonable to formulate the component problem in the sequence as a Bayes decision problem with respect to an unknown prior distribution on the parameter space. One then uses the accumulated observations to improve the decision rule at each stage. This is the empirical Bayes approach of Robbins (1956,1964). During the last thirty years, empirical Bayes methods have been studied extensively. Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the n^{th} component decision problem converges to the minimum Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used.

Empirical Bayes rules have been derived for subset selection goals by Deely (1965). Recently, Gupta and Liang (1986,1988a,1988b) have studied empirical Bayes rules for selecting binomial and negative binomial populations better than a standard or a control and for selecting the best among several binomial populations. They have assumed that

the form of the prior distributions are completely unknown. Hence, those approaches are referred to as nonparametric empirical Bayes. Gupta and Liang (1989a,1989b) have also studied some other empirical Bayes selection rules, in which they assumed that the form of the prior distribution is known but the distributions depend on certain unknown hyperparameters. Such approach is therefore referred to as parametric empirical Bayes.

In this paper, we are concerned with the problem of selecting the highest probability of success among several negative binomial distributions through the nonparametric empirical Bayes approach. The selection rule is based on monotone empirical Bayes estimators of the negative binomial success probabilities. The framework of the empirical Bayes selection problem is formulated in Section 2. Using the isotonic regression method, monotone empirical Bayes estimators as well as a monotone empirical Bayes selection rule are proposed in Section 3. Finally, the asymptotic optimality property of the proposed empirical Bayes selection rule is studied in Section 4.

2. FORMULATION OF THE EMPIRICAL BAYES APPROACH

Consider $k(\geq 2)$ independent negative binomial populations $\pi_1, \ldots, \pi_k$. For each $i = 1, \ldots, k$, let p_i denote the probability of success for each trial in π_i and let X_i denote the number of successes before attaining the r^{th} failure in π_i. We assume that for each $i = 1, \ldots, k$, the trials in π_i are mutually independent. Thus, conditional on p_i, $X_i|p_i$ has a negative binomial distribution with probability function $f_i(x|p_i)$, where

$$f_i(x|p_i) = \binom{x+r-1}{r-1} p_i^x (1-p_i)^r, \; x = 0, 1, 2, \ldots. \tag{2.1}$$

Let $f(\underline{x}|\underline{p}) = \prod_{i=1}^{k} f_i(x_i|p_i)$, where $\underline{x} = (x_1, \ldots, x_k)$ and $\underline{p} = (p_1, \ldots, p_k)$. For each $\underline{p}$, let

$p_{[1]} \leq \cdots \leq p_{[k]}$ denote the ordered values of the parameters $p_1, \ldots, p_k$. It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. A population π_i with $p_i = p_{[k]}$ is referred to as a best population. Our goal is to derive empirical Bayes rules to select the best population. The empirical Bayes framework of the selection problem is formulated as follows.

Let $\Omega = \{\underline{p} = (p_1, \ldots, p_k) | p_i \in (0,1),\ i = 1, \ldots, k\}$ be the parameter space and let $G(\underline{p}) = \prod_{i=1}^{k} G_i(p_i)$ be the prior distribution over Ω, where $G_i(\cdot)$ are unknown for all $i = 1, \ldots, k$. Note that under this model, p_i's are assumed to be independently distributed. Let $A = \{i | i = 1, \ldots, k\}$ be the action space. When action i is taken, it means that population π_i is selected as the best population. For the parameter $\underline{p}$ and the action i, the loss function $L(\underline{p}, i)$ is defined as:

$$L(\underline{p}, i) = p_{[k]} - p_i, \tag{2.2}$$

the difference between the best and the selected population.

Let $\mathcal{X}$ be the sample space generated by $\underline{X} = (X_1, \ldots, X_k)$. A selection rule $d = (d_1, \ldots, d_k)$ is defined to be a mapping from the sample space $\mathcal{X}$ to $[0,1]^k$ such that for each observation $\underline{x} = (x_1, \ldots, x_k)$, the function $d(\underline{x}) = (d_1(\underline{x}), \ldots, d_k(\underline{x}))$ satisfies that $0 \leq d_i(\underline{x}) \leq 1$ for all $i = 1, \ldots, k$, and $\sum_{i=1}^{k} d_i(\underline{x}) = 1$. Note that $d_i(\underline{x})$ is the probability of selecting population π_i as the best population when $\underline{x}$ is observed. Let D be the set of all selection rules defined previously. For each $d \in D$, let $r(G, d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d \in D} r(G, d)$ is the minimum Bayes risk among the class D, and a rule, say d_G, is called a Bayes selection rule if $r(G, d_G) = r(G)$.

Based on the preceding statistical model, the Bayes risk associated with the selection rule d is:

$$\begin{aligned} r(G,d) &= \int_{\Omega} \sum_{\underline{x}\in \mathcal{X}} \sum_{i=1}^{k} L(\underline{p},i) d_i(\underline{x}) f(\underline{x}|\underline{p}) dG(\underline{p}) \\ &= C - \sum_{\underline{x}\in\mathcal{X}} \left[\sum_{i=1}^{k} d_i(\underline{x}) \varphi_i(x_i) \right] f(\underline{x}) \end{aligned} \tag{2.3}$$

where

$$\begin{cases} f(\underline{x}) &= \prod_{i=1}^{k} f_i(x_i), \\ f_i(x_i) &= \int_0^1 f_i(x_i|p) dG_i(p) = \binom{x_i+r-1}{r-1} \int_0^1 p^{x_i}(1-p)^r dG_i(p) = a(x_i) h_i(x_i), \\ a(x_i) &= \binom{x_i+r-1}{r-1}, \ h_i(x_i) = \int_0^1 p^{x_i}(1-p)^r dG_i(p), \\ \varphi_i(x_i) &= h_i(x_i+1)/h_i(x_i) \ \text{(note that } 0 < \varphi_i(x_i) < 1\text{), and} \\ C &= \sum_{\underline{x}\in\mathcal{X}} \int_{\Omega} p_{[k]} f(\underline{x}|\underline{p}) dG(\underline{p}). \end{cases} \tag{2.4}$$

Note that C is a constant which is independent of the selection rule d. Thus, from (2.3), Bayes selection rules can be obtained as follows.

For each $\underline{x} \in \mathcal{X}$, let

$$A(\underline{x}) = \{ i | \varphi_i(x_i) = \max_{i \le j \le k} \varphi_j(x_j) \}. \tag{2.5}$$

Any selection rule $d = (d_1, \ldots, d_k)$ such that $\sum_{i\in A(\underline{x})} d_i(\underline{x}) = 1$ is always a Bayes rule. Thus, a randomized Bayes rule is $d_G = (d_{1G}, \ldots, d_{kG})$ where

$$d_{iG}(\underline{x}) = \begin{cases} |A(\underline{x})|^{-1} & \text{if } i \in A(\underline{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

and $|A|$ denotes the cardinality of the set A.

Since the prior distribution G is unknown, it is not possible to apply the Bayes rules for the selection problem at hand. In this case, we use the empirical Bayes approach. It is assumed that past observations from each of the k populations are available.

For each $i = 1, \ldots, k$, let (X_{ij}, P_{ij}), $j = 1, \ldots, n$, be independent random vectors associated with population π_i, where P_{ij} stands for the random probability of success for each trial in π_i at stage j, and X_{ij} stands for the number of successes before attaining the r^{th} failure in π_i at stage j. It is assumed that P_{ij} has prior distribution G_i for all $j = 1, 2, \ldots$. Conditional on $P_{ij} = p_{ij}$, $X_{ij}|p_{ij}$ has a negative binomial probability function $f_i(x_{ij}|p_{ij})$ given in (2.1). It should be noted that X_{ij} is observable but P_{ij} is not. Let the j^{th} stage observations be denoted by $\underline{X}_j$. That is, $\underline{X}_j = (X_{1j}, \ldots, X_{kj})$. From the assumptions, $\underline{X}_1, \ldots, \underline{X}_n$ are mutually independent and identically distributed. We also let $\underline{X}_{n+1} = \underline{X} = (X_1, \ldots, X_k)$ denote the observation at the current stage.

From (2.5) and (2.6), a natural empirical Bayes selection rule can be defined as follows: For each $i = 1, \ldots, k$, and $n = 1, 2, \ldots$, based on the past data $X_{i1}, \ldots, X_{in}$ and the present observation $X_i = x_i$, let $\varphi_{in}(x_i) = \varphi_{in}(x_i; X_{i1}, \ldots, X_{in})$ be an empirical Bayes estimator of $\varphi_i(x_i)$. Then, by letting

$$A_n(\underline{x}) = \{i | \varphi_{in}(x_i) = \max_{1 \le j \le k} \varphi_{jn}(x_j)\}, \tag{2.7}$$

an empirical Bayes selection rule $d_n(\underline{x}) = (d_{1n}(\underline{x}), \ldots, d_{kn}(\underline{x}))$ is defined as below.

$$d_{in}(\underline{x}) = \begin{cases} |A_n(\underline{x})|^{-1} & \text{if } i \in A_n(\underline{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

Note that the past data $(\underline{X}_1, \ldots, \underline{X}_n)$ is implicitly contained in the subscript n.

For such an empirical Bayes selection rule d_n, let $r(G, d_n)$ be the corresponding overall Bayes risk. That is,

$$
\begin{aligned}
r(G, d_n) &= E\left[\int_\Omega \sum_{\underline{x}\in\mathcal{X}} \sum_{i=1}^{k} L(\underline{p}, i) d_{in}(\underline{x}) f(\underline{x}|\underline{p}) dG(\underline{p})\right] \\
&= C - \sum_{\underline{x}\in\mathcal{X}} \left[\sum_{i=1}^{k} E[d_{in}(\underline{x})]\varphi_i(x_i)\right] f(\underline{x}),
\end{aligned}
\tag{2.9}
$$

where the expectation E is taken with respect to $(\underline{X}_1, \ldots, \underline{X}_n)$. Since $r(G)$ is the minimum Bayes risk, $r(G, d_n) - r(G) \geq 0$ for all n. Thus, the nonnegative difference $r(G, d_n) - r(G)$ is used as a measure of the optimality of the empirical Bayes selection rule d_n.

Definition 2.1. A sequence of empirical Bayes selection rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the (unknown) prior distribution G if $r(G, d_n) - r(G) \to 0$ as $n \to \infty$.

In the following, we seek a sequence of asymptotically optimal empirical Bayes selection rules for the selection problem under study.

3. THE PROPOSED EMPIRICAL BAYES SELECTION RULE

Before we go further to construct empirical Bayes rules for the selection problem at hand, we first investigate some properties related to the Bayes selection rule d_G defined in (2.6).

Definition 3.1. A selection rule $d = (d_1, \ldots, d_k)$ is said to be monotone if for each $i = 1, \ldots, k$, $d_i(\underline{x})$ is increasing in x_i while all other variables x_j are fixed, and decreasing in x_j, for each $j \neq i$ while all other variables are fixed.

Note that for each $i = 1, \ldots, k$, $\varphi_i(x_i) = h_i(x_i + 1)/h_i(x_i)$. Straight computations

show that $\varphi_i(x_i)$ is an increasing function of x_i. Thus, from (2.5) and (2.6), one can see that the Bayes selection rule d_G is a monotone selection rule. Also, note that $\varphi_i(x_i)$ is the posterior mean of P_i given $X_i = x_i$, and it is the Bayes estimator of P_i given $X_i = x_i$ for squared error loss. Under the squared error loss, the problem of estimating the probability of success in a negative binomial distribution is a monotone estimation problem. By Theorem 8.7 of Berger (1985), for a monotone estimation problem, the class of monotone estimators form an essential complete class. Also, for the present selection problem, under the loss function given in (2.2), the problem is a monotone decision problem. Again, from Berger (1985), the class of monotone selection rules is essentially complete. Now, one can see that if the empirical Bayes estimators $\varphi_{in}(x_i)$, $i = 1, \ldots, k$, are monotone, then the empirical Bayes selection rule given through (2.7) and (2.8) is also monotone. From these considerations, it is reasonable to desire that the concerned estimators $\{\varphi_{in}(x_i)\}$ possess the above–mentioned monotonicity property.

For each $i = 1, \ldots, k$, let $N_{in} = \max(X_{i1}, \ldots, X_{in}) - 1$. For each $x = 0, 1, 2, \ldots$, let

$$f_{in}(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_{ij}), \tag{3.1}$$

where I_A denotes the indicator function oi the set A. Let

$$h_{in}(x) = f_{in}(x)/a(x). \tag{3.2}$$

It is intuitive to use $h_{in}(x)$ as an estimator of $h_i(x)$. However, note that $h_i(x)$ is nonincreasing in x; while $h_{in}(x)$ may not possess the nonincreasing property. Thus, we consider a smoothed version of $h_{in}(x)$. Note that $h_{in}(x) = 0$ if $x > N_{in}+1$. Thus, let $\{h^*_{in}(x)\}_{x=0}^{N_{in}+1}$ be the antitonic regression of $\{h_{in}(x)\}_{x=0}^{N_{in}+1}$ with equal weights, and let $h^*_{in}(y) = 0$ for all $y > N_{in} + 1$. Thus, $h^*_{in}(x)$ is nonincreasing in x.

For $x = 0, 1, \ldots, N_{in}$, let

$$\varphi_{in}(x) = h^*_{in}(x+1)/h^*_{in}(x). \tag{3.3}$$

Since $\varphi_{in}(x)$ may not possess the nondecreasing property as $\varphi_i(x)$ does, we consider a smoothed version of $\varphi_{in}(x)$ as follows. Let $\{\varphi^*_{in}(x)\}_{x=0}^{N_{in}}$ be the isotonic regression of $\{\varphi_{in}(x)\}_{x=0}^{N_{in}}$ with random weights $\{h^*_{in}(x)\}_{x=0}^{N_{in}}$. For $y > N_{in}$, we define $\varphi^*_{in}(y) = \varphi^*_{in}(N_{in})$. Therefore $\varphi^*_{in}(x)$ is nondecreasing in x. Note that $0 \leq \varphi^*_{in}(x) \leq 1$ for all $x = 0, 1, 2, \ldots$. We use $\varphi^*_{in}(x)$ to estimate $\varphi_i(x)$ and propose an empirical Bayes selection rule $d^*_n = (d^*_{1n}, \ldots, d^*_{kn})$ as follows.

For each $\underline{x} = (x_1, \ldots, x_k) \in \mathcal{X}$, let

$$A^*_n(\underline{x}) = \{i | \varphi^*_{in}(x_i) = \max_{1 \leq j \leq k} \varphi^*_{jn}(x_j)\} \tag{3.4}$$

and for each $i = 1, \ldots, k$, let

$$d^*_{in}(\underline{x}) = \begin{cases} |A^*_n(\underline{x})|^{-1} & \text{if } i \in A^*_n(\underline{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Since $\varphi^*_{in}(x_i)$ is nondecreasing in x_i for each $i = 1, \ldots, k$, we see that $d^*_n(\underline{x})$ is a monotone empirical Bayes selection rule.

<u>Remark 3.1</u>. For each $y = 0, 1, 2, \ldots$, let $H_{in}(y) = \sum_{x=0}^{y} h_{in}(x)$, $H^*_{in}(y) = \sum_{x=0}^{y} h^*_{in}(x)$, and $H_i(y) = \sum_{x=0}^{y} h_i(x)$. Also, for each $y = 0, 1, \ldots, N_{in}$, let $\Psi_{in}(y) = \sum_{x=0}^{y} \varphi_{in}(x) h^*_{in}(x)$, and $\Psi^*_{in}(y) = \sum_{x=0}^{y} \varphi^*_{in}(x) h^*_{in}(x)$. Then, by the definition of $\varphi_{in}(x)$, $\psi_{in}(y) = \sum_{x=0}^{y} h^*_{in}(x+1) = H^*_{in}(y+1) - H^*_{in}(0)$ for $0 \leq y \leq N_{in}$. Also, from Barlow, et al. (1972), and by noting the fact that $h_{in}(y) = h^*_{in}(y) = 0$ for all $y > N_{in} + 1$, we have the following results:

$$H^*_{in}(y) \geq H_{in}(y) \text{ for all } y = 0, 1, 2, \ldots \tag{3.6}$$

$$\sup_{y\ge 0}|H^*_{in}(y) - H_i(y)| \le \sup_{y\ge 0}|H_{in}(y) - H_i(y)|. \tag{3.7}$$

$$\Psi^*_{in}(y) \le \Psi_{in}(y) \text{ for all } y = 0, 1, \ldots \tag{3.8}$$

Remark 3.2. From Puri and Singh (1988), we have

$$\varphi^*_{in}(0) = \min_{0\le y\le N_{in}} [\Psi_{in}(y)/H^*_{in}(y)]. \tag{3.9}$$

$$\varphi^*_{in}(x) = \min_{x\le y\le N_{in}} [(\Psi_{in}(y) - \Psi^*_{in}(x-1))/(H^*_{in}(y) - H^*_{in}(x-1))], \; x = 1, \ldots, N_{in}. \tag{3.10}$$

From (3.8), (3.9) and (3.10) we obtain

$$\begin{aligned} \varphi^*_{in}(x) &\ge \min_{x\le y\le N_{in}} [(\Psi_{in}(y) - \Psi_{in}(x-1))/(H^*_{in}(y) - H^*_{in}(x-1))] \\ &= \min_{x\le y\le N_{in}} [(H^*_{in}(y+1) - H^*_{in}(x))/(H^*_{in}(y) - H^*_{in}(x-1))] \end{aligned} \tag{3.11}$$

for each $x = 0, 1, \ldots, N_{in}$, where $H^*_{in}(-1) \equiv 0$.

Remark 3.3. Analogous to Puri and Singh (1988), we can obtain an alternative form of $\varphi^*_{in}(x)$ as follows.

$$\varphi^*_{in}(N_{in}) = \max_{0\le y\le N_{in}} \left[\left(\sum_{r=y}^{N_{in}} \varphi_{in}(r) h^*_{in}(r) \right) / \sum_{r=y}^{N_{in}} h^*_{in}(r) \right] \tag{3.12}$$

$$\varphi^*_{in}(x) = \max_{0\le y\le x} \left[\left(\sum_{r=y}^{N_{in}} \varphi_{in}(r) h^*_{in}(r) - \sum_{r=x+1}^{N_{in}} \varphi^*_{in}(r) h^*_{in}(r) \right) / \sum_{r=y}^{x} h^*_{in}(r) \right] \tag{3.13}$$

for $x = 0, 1, \ldots, N_{in} - 1$, and

$$\sum_{r=x}^{N_{in}} \varphi_{in}(r) h^*_{in}(r) \le \sum_{r=x}^{N_{in}} \varphi^*_{in}(r) h^*_{in}(r), \quad \text{for } x = 0, 1, \ldots, N_{in}. \tag{3.14}$$

Thus, we can obtain

$$\begin{aligned} \varphi^*_{in}(x) &\le \max_{0\le y\le x} \left[\left(\sum_{r=y}^{N_{in}} \varphi_{in}(r) h^*_{in}(r) - \sum_{r=x+1}^{N_{in}} \varphi_{in}(r) h^*_{in}(r) \right) / \sum_{r=y}^{x} h^*_{in}(r) \right] \\ &= \max_{0\le y\le x} [(H^*_{in}(x+1) - H^*_{in}(y)) \,/\, (H^*_{in}(x) - H^*_{in}(y-1))] \end{aligned} \tag{3.15}$$

for each $x = 0, 1, \ldots, N_{in}$.

4. ASYMPTOTIC OPTIMALITY

In this section, we investigate the asymptotic optimality property of the sequence of empirical Bayes selection rules $\{d_n^*\}$ defined previously.

For $1 \le i < j \le k$, and each $x_i = 0, 1, 2 \ldots$, let $A_{ij}(x_i) = \{x_j | \varphi_j(x_j) > \varphi_i(x_i)\}$ and $B_{ij}(x_i) = \{x_j | \varphi_j(x_j) < \varphi_i(x_i)\}$. Let

$$M_{ij}(x_i) = \begin{cases} \min\ A_{ij}(x_i) & \text{if } A_{ij}(x_i) \neq \phi, \\ \infty & \text{if } A_{ij}(x_i) = \phi, \end{cases}$$

$$m_{ij}(x_i) = \begin{cases} \max\ B_{ij}(x_i) & \text{if } B_{ij}(x_i) \neq \phi, \\ -1 & \text{if } B_{ij}(x_i) = \phi. \end{cases}$$

Note that, by the nondecreasing property of $\varphi_j(x_j)$, $m_{ij}(x_i) \le M_{ij}(x_i)$ and $m_{ij}(x_i) < M_{ij}(x_i)$ if $A_{ij}(x_i) \neq \phi$. Let

$$\begin{aligned} r_{ij}(n) = &\sum_{x_i=0}^{\infty} \sum_{x_j=0}^{m_{ij}(x_i)} [\varphi_i(x_i) - \varphi_j(x_j)] P\{\varphi_{in}^*(x_i) \le \varphi_{jn}^*(x_j)\} f_i(x_i) f_j(x_j) \\ &+ \sum_{x_i=0}^{\infty} \sum_{x_j=M_{ij}(x_i)}^{\infty} [\varphi_j(x_j) - \varphi_i(x_i)] P\{\varphi_{in}^*(x_i) \ge \varphi_{jn}^*(x_j)\} f_i(x_i) f_j(x_j) \end{aligned}$$

where $\sum_{x_j=0}^{m_{ij}(x_i)} \equiv 0$ if $m_{ij}(x_i) = -1$ and $\sum_{x_j=M_{ij}(x_i)}^{\infty} \equiv 0$ if $M_{ij}(x_i) = \infty$.

From (2.5) and (2.9), a straightforward computation leads to that

$$\begin{aligned} 0 &\le r(G, d_n^*) - r(G) \\ &\le \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} r_{ij}(n). \end{aligned}$$

Thus, in order to investigate the asymptotic behavior of the difference $r(G, d_n^*) - r(G)$, it suffices to consider the case where $k = 2$. We claim that

<u>Theorem 4.1</u> $r_{12}(n) \to 0$ as $n \to \infty$.

Theorem 4.1 implies that the sequence of empirical Bayes selection rules $\{d_n^*\}$ is asymptotically optimal. In the rest of this section, we are going to prove Theorem 4.1. The following lemmas are useful in presenting a concise proof of Theorem 4.1.

<u>Lemma 4.1</u> Let $\{a_m\}$ be a sequence of real numbers and $\{b_m\}$ be a sequence of positive numbers such that $b_m \leq 1$ and b_m is nonincreasing in m. Then, for each positive constant c,

$$\sup_{n\geq 1} |\sum_{m=1}^{n} a_m b_m| \geq (>) c \text{ implies that } \sup_{n\geq 1} |\sum_{m=1}^{n} a_m| \geq (>) c.$$

Since this lemma is trivial, the proof is omitted here. The following is a consequent result of Lemma 4.1.

<u>Corollary 4.1</u> For each $i = 1, \ldots, k$, and $z = 0, 1, 2 \ldots$, let $F_i(z) = \sum_{x=0}^{z} f_i(x)$, $F_{in}(z) = \sum_{x=0}^{z} f_{in}(x)$. Then, for fixed positive constant c,

$$\sup_{z\geq 0} |H_{in}(z) - H_i(z)| \geq (>) c \text{ implies that } \sup_{z\geq 0} |F_{in}(z) - F_i(z)| \geq (>) c.$$

Proof: There is a direct result of Lemma 4.1 by noting that $h_{in}(x) - h_i(x) = (f_{in}(x) - f_i(x))/a(x)$, where $1/a(x) = \binom{x+r-1}{r-1}^{-1} \leq 1$ and which is decreasing in x.

<u>Lemma 4.2</u>. For each fixed $x = 0, 1, 2 \ldots$, and $0 < t < \varphi_i(x)$, let $Q_i(y|x,t) = [-H_i(y+1) + H_i(x)] + [H_i(y) - H_i(x-1)][\varphi_i(x) - t]$ for all $y = x, x+1, \ldots$. Then, $Q_i(y|x,t)$ is nonincreasing in y and therefore, $\max_{y\geq x} Q_i(y|x,t) = Q_i(x|x,t) = -th_i(x) < 0$.

Proof: $Q_i(y|x,t) - Q_i(y+1|x,t)$

$$= [H_i(y+2) - H_i(y+1)] + [H_i(y) - H_i(y+1)][\varphi_i(x) - t]$$

$$= h_i(y+2) - h_i(y+1)[\varphi_i(x) - t]$$

$$= h_i(y+1)[\varphi_i(y+1) - \varphi_i(x) + t]$$

$$\geq h_i(y+1)t$$

$$> 0$$

where the first inequality is due to the fact that $y \geq x$ and thus $\varphi_i(y+1) \geq \varphi_i(x)$. Therefore, $\varphi_i(y|x,t)$ is nonincreasing in y for all $y \geq x$, and hence $\max\limits_{y \geq x} Q_i(y|x,t) = Q_i(x|x,t) = -th_i(x) < 0$.

<u>Lemma 4.3</u>. For each $x = 0,1,2,\ldots$, and $0 < t < 1 - \varphi_i(x)$, let $R_i(y|x,t) = [-H_i(x+1) + H_i(y)] + [H_i(x) - H_i(y-1)]\ [\varphi_i(x) + t]$ for all $y = 0,1,\ldots,x$, where $H_i(-1) \equiv 0$. Then, $R_i(y|x,t)$ is nonincreasing in y for $y = 0,1,\ldots,x$, and hence, $\min\limits_{0 \leq y \leq x} R_i(y|x,t) = R_i(x|x,t) = th_i(x) > 0$.

Proof: For $0 \leq y \leq x - 1$,

$$R_i(y|x,t) - R_i(y+1|x,t)$$

$$= [H_i(y) - H_i(y+1)] + [H_i(y) - H_i(y-1)][\varphi_i(x) + t]$$

$$= -h_i(y+1) + h_i(y)[\varphi_i(x) + t]$$

$$= h_i(y)[-\varphi_i(y) + \varphi_i(x) + t]$$

$$\geq h_i(y)t$$

$$> 0$$

since $0 \leq y \leq x - 1$ and thus $\varphi_i(x) - \varphi_i(y) \geq 0$. Therefore, $R_i(y|x,t)$ is nonincreasing in y for $y = 0,1,\ldots,x$, and hence $\min\limits_{0 \leq y \leq x} R_i(y|x,t) = R_i(x|x,t) = h_i(x)t > 0$.

Lemma 4.4. For each $i = 1, \ldots, k,\ t > 0$ and $x = 0, 1, 2, \ldots$

$$P\{\varphi_{in}^*(x) - \varphi_i(x) \le -t\} \le [F_i(x)]^n + d\exp(-nt^2 h_i^2(x)/8)$$

where d is a positive constant which is independent of the distribution F_i.

Proof: Note that $0 < \varphi_{in}^*(x),\ \varphi_i(x) < 1$. Hence $P\{\varphi_{in}^*(x) - \varphi_i(x) \le -t\} = 0$ if $t \ge \varphi_i(x)$. Thus, in the following, it suffices to consider those t such that $0 < t < \varphi_i(x)$.

Now, for $0 < t < \varphi_i(x)$,

$$\begin{aligned} &P\{\varphi_{in}^*(x) - \varphi_i(x) \le -t\} \\ =&P\{\varphi_{in}^*(x) - \varphi_i(x) \le -t \text{ and } N_{in} < x\} + P\{\varphi_{in}^*(x) - \varphi_i(x) \le -t \text{ and } N_{in} \ge x\}, \end{aligned}$$

where

$$P\{\varphi_{in}^*(x) - \varphi_i(x) \le -t \text{ and } N_{in} < x\} \le [F_i(x)]^n \tag{4.2}$$

which is obtained by the definition of N_{in}.

Let $T_{in}^*(x) = H_{in}^*(x) - H_i(x)$ and $T_{in}(x) = H_{in}(x) - H_i(x)$. From (3.11), (3.7), Corollary 4.1, Lemma 4.2 and Lemma 2.1 of Schuster (1969) and by noting the fact that $0 < t < \varphi_i(x) < 1$, we have

$$
\begin{aligned}
&P\{\varphi^*_{in}(x) - \varphi_i(x) \leq -t \text{ and } N_{in} \geq x\}\\
\leq &P\{H^*_{in}(y+1) - H^*_{in}(x) - [H^*_{in}(y) - H^*_{in}(x-1)][\varphi_i(x) - t] \leq 0 \text{ for some } x \leq y \leq N_{in}\}\\
= &P\{T^*_{in}(y+1) - T^*_{in}(x) - [T^*_{in}(y) - T^*_{in}(x-1)][\varphi_i(x) - t] \leq Q_i(y|x,t) \text{ for some } x \leq y \leq N_{in}\}\\
\leq &P\{T^*_{in}(y+1) - T^*_{in}(x) - [T^*_{in}(y) - T^*_{in}(x-1)][\varphi_i(x) - t] \leq -th_i(x) \text{ for some } x \leq y \leq N_{in}\}\\
\leq &P\{\sup_{x \leq z \leq N_{in}+1} |T^*_{in}(z)| > th_i(x)/4\} \qquad (4.3)\\
\leq &P\{\sup_{z \geq 0} |T^*_{in}(z)| > th_i(x)/4\}\\
\leq &P\{\sup_{z \geq 0} |T_{in}(z)| > th_i(x)/4\}\\
\leq &P\{\sup_{z \geq 0} |F_{in}(z) - F_i(z)| > th_i(x)/4\}\\
\leq &d \exp(-nt^2 h_i^2(x)/8).
\end{aligned}
$$

Thus, from (4.1), (4.2) and (4.3), the proof of this lemma is complete.

Lemma 4.5. For each $i = 1, \ldots, k,\ t > 0$ and $x = 0, 1, 2, \ldots,$

$$P\{\varphi^*_{in}(x) - \varphi_i(x) \geq t\} \leq [F_i(x)]^n + d\exp(-nt^2 h_i^2(x)/8)$$

where d is a positive constant which is independent of the distribution F_i.

Proof: The proof of this lemma uses the results of (3.7), (3.14), Corollary 4.1, Lemma 4.3 and Lemma 2.1 of Schuster (1969). The argument for the proof of this lemma is similar to that of Lemma 4.4 We omit the detail here.

Proof of Theorem 4.1

Let $\Delta(x_1, x_2) = \varphi_1(x_1) - \varphi_2(x_2)$. Note that

$$\begin{aligned} r_{12}(n) = &\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{m_{12}(x_1)} \Delta(x_1, x_2) P\{\varphi_{1n}^*(x_1) \le \varphi_{2n}^*(x_2)\} f_1(x_1) f_2(x_2) \\ &+ \sum_{x_1=0}^{\infty} \sum_{x_2=M_{12}(x_1)}^{\infty} (-\Delta(x_1, x_2)) P\{\varphi_{1n}^*(x_1) \ge \varphi_{2n}^*(x_2)\} f_1(x_1) f_2(x_2). \end{aligned} \tag{4.4}$$

For $x_2 \le m_{12}(x_1)$, $\varphi_1(x_1) - \varphi_2(x_2) > 0$ and

$$\begin{aligned} &P\{\varphi_{1n}^*(x_1) \le \varphi_{2n}^*(x_2)\} \\ \le &P\left\{\varphi_{1n}^*(x_1) - \varphi_1(x_1) \le -\frac{\varphi_1(x_1) - \varphi_2(x_2)}{2}\right\} \\ &+ P\left\{\varphi_{2n}^*(x_2) - \varphi_2(x_2) \ge \frac{\varphi_1(x_1) - \varphi_2(x_2)}{2}\right\}. \end{aligned} \tag{4.5}$$

For $x_2 \ge M_{12}(x_1)$, $\varphi_1(x_1) - \varphi_2(x_2) < 0$ and,

$$\begin{aligned} &P\{\varphi_{1n}^*(x_1) \ge \varphi_{2n}^*(x_2)\} \\ \le &P\left\{\varphi_{1n}^*(x_1) - \varphi_1(x_1) \ge \frac{\varphi_2(x_2) - \varphi_1(x_1)}{2}\right\} \\ &+ P\left\{\varphi_{2n}^*(x_2) - \varphi_2(x_2) \le -\frac{\varphi_2(x_2) - \varphi_1(x_1)}{2}\right\}. \end{aligned} \tag{4.6}$$

Also, $0 < \frac{\varphi_1(x_1) - \varphi_2(x_2)}{2} < \min(\varphi_1(x_1),\ 1 - \varphi_2(x_2))$ as $x_2 \le m_{12}(x_1)$, and $0 < \frac{\varphi_2(x_2) - \varphi_1(x_1)}{2} < \min(\varphi_2(x_2),\ 1 - \varphi_1(x_1))$ as $x_2 \ge M_{12}(x_1)$. Then, from (4.4)–(4.6), Lemmas 4.4 and 4.5, one can obtain the following:

$$\begin{aligned} r_{12}(n) \le &E[F_1^n(X_1)] + E[F_2^n(X_2)] \\ &+ d \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_1(x_1) f_2(x_2) \left[\exp(-n\Delta^2(x_1, x_2) h_1^2(x_1)/8) + \exp(-n\Delta^2(x_1, x_2) h_2^2(x_2)/8)\right] \end{aligned} \tag{4.7}$$

where each of the three terms at the right-hand-side of (4.7) tends to 0 as n tends to infinity. This implies that $r_{12}(n) \to 0$ as $n \to \infty$. Hence the proof of the theorem is complete.

Acknowledgement: The author is thankful to Professor Shanti S. Gupta for encouragement and for support of this research. This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620 at Purdue University.

BIBLIOGRAPHY

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Deely, J.J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo. Ser. No. 45), Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S. and Liang, T. (1986). Empirical Bayes rules for selecting good binomial populations. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes-Monograph Series, Vol. 8, 110-128.

Gupta, S.S. and Liang, T. (1988a). On empirical Bayes selection rules for negative binomial populations. Technical Report #88-16C, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S. and Liang, T. (1988b). Empirical Bayes rules for selecting the best binomial population. *Statistical Decision Theory and Related Topics-IV* (Eds. S.S. Gupta and J.O. Berger), Springer-Verlag, Vol. I, 213-224.

Gupta, S.S. and Liang, T. (1989a). Parametric empirical Bayes rules for selecting the most probable multinomial event. To appear in a Volume in honor of Ingram Olkin.

Gupta, S.S. and Liang, T. (1989b). Selecting the best binomial population: parametric empirical Bayes approach. To appear in *J. Statist. Plann. Inference*.

Gupta, S.S. and Panchapakesan, S. (1979). *Multiple Decision Procedures*. Wiley, New York.

Puri, P.S. and Singh, H. (1988). On recursive formulas for isotonic regression useful for statistical inference under order restrictions. Technical Report #88–51, Department of Statistics, Purdue University, West Lafayette, Indiana.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, University of California Press, 157–163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1–20.

Schuster, E.F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40**, 1187–1195.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #89-18C	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 47907	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170 NSF DMS-8606964, DMS-8702620
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO. | WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
EMPIRICAL BAYES SELECTION FOR THE HIGHEST PROBABILITY OF SUCCESS IN NEGATIVE BINOMIAL DISTRIBUTIONS

12. PERSONAL AUTHOR(S)
TaChen Liang

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ________ TO ________	June 1989	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Asymptotically optimal; empirical Bayes; monotone estimator; monotone selection rule; isotonic and antitonic regression.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We study the problem of selecting the highest probability of success from among several negative binomial distributions via the nonparametric empirical Bayes approach. A monotone selection rule is proposed on basis of monotone empirical Bayes estimators of the negative binomial success probabilities which are obtained by using the antitonic and isotonic regression methods. The asymptotic optimality property of the proposed empirical Bayes selection rule is also established.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Shanti S. Gupta	22b. TELEPHONE (Include Area Code) 317-494-6031	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED